

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
[2]: # Import all the libraries that we need.
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
[3]: df = pd.read_csv('C:\\Users\\siva\\Downloads\\heart.csv')
```

```
[4]: df.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

```
[5]: df.tail()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

```
[6]: df.columns.values
```

```
[6]: array(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg',
       'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
      dtype=object)
```

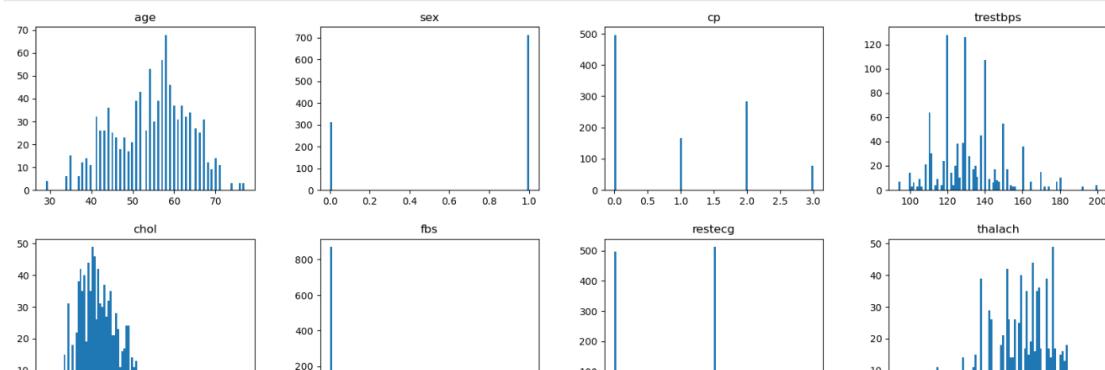
```
[7]: df.isna().sum()
```

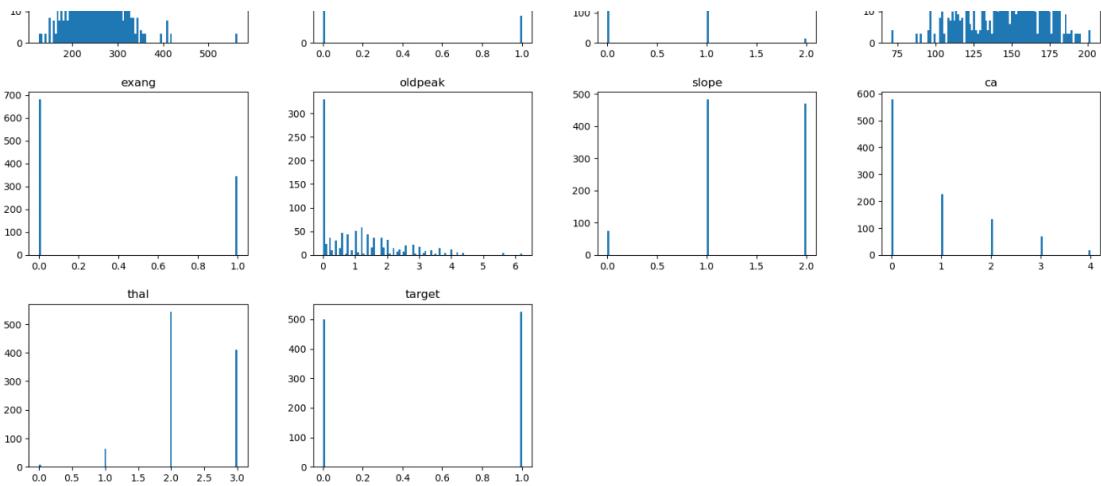
```
[7]: age      0
sex      0
cp      0
trestbps  0
chol      0
fbs      0
restecg    0
thalach    0
exang      0
oldpeak    0
slope      0
ca      0
thal      0
target      0
dtype: int64
```

```
[8]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   age      1025 non-null   int64  
 1   sex      1025 non-null   int64  
 2   cp       1025 non-null   int64  
 3   trestbps 1025 non-null   int64  
 4   chol     1025 non-null   int64  
 5   fbs      1025 non-null   int64  
 6   restecg  1025 non-null   int64  
 7   thalach  1025 non-null   int64  
 8   exang    1025 non-null   int64  
 9   oldpeak  1025 non-null   float64 
 10  slope    1025 non-null   int64  
 11  ca       1025 non-null   int64  
 12  thal    1025 non-null   int64  
 13  target   1025 non-null   int64  
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
```

```
[9]: df.hist(bins=100, grid = False, figsize =(20,15));
```





```
[10]: df.describe()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695610	0.942439	131.611707	246.000000	0.149268	0.529756	149.114146	0.336585	1.071512	1.385366	0.754146
std	9.072290	0.460373	1.029641	17.516718	51.59251	0.356527	0.527878	23.005724	0.472772	1.175053	0.617755	1.030798
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000
25%	48.000000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000
50%	56.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	152.000000	0.000000	0.800000	1.000000	0.000000
75%	61.000000	1.000000	2.000000	140.000000	275.000000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000

```
[13]: questions = ["1. How many people have heart disease and how many people doesn't have heart disease?",  
                 "2. People of which sex has most heart disease?",  
                 "3. People of which sex has which type of chest pain most?",  
                 "4. People with which chest pain are most prone to have heart disease?",  
                 "5. People of what age has most heart disease?",  
                 "6. People of what age has most chest pain?"]
```

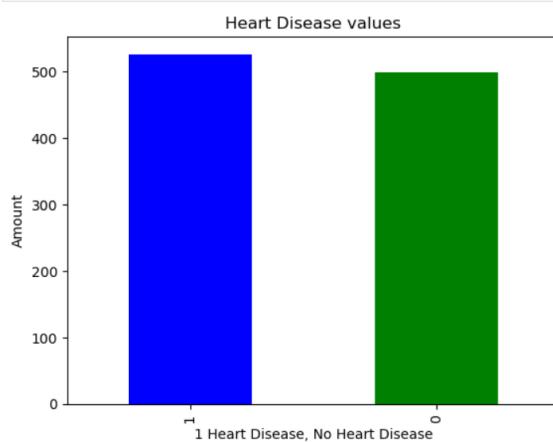
questions

```
[13]: [{"1. How many people have heart disease and how many people doesn't have heart disease?",  
   '2. People of which sex has most heart disease?',  
   '3. People of which sex has which type of chest pain most?',  
   '4. People with which chest pain are most prone to have heart disease?',  
   '5. People of what age has most heart disease?',  
   '6. People of what age has most chest pain?']}
```

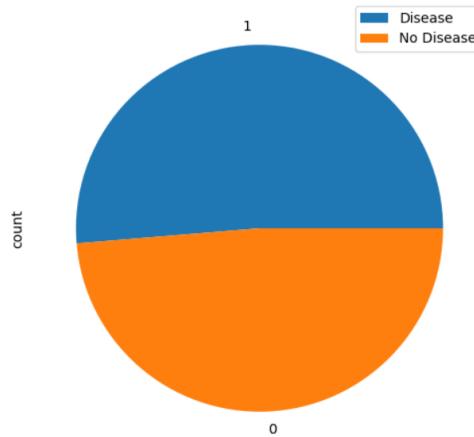
```
[15]: #question 1  
df.target.value_counts()
```

```
[15]: target  
1    526  
0    499  
Name: count, dtype: int64
```

```
[21]: import matplotlib.pyplot as plt  
  
df.target.value_counts().plot(kind='bar', color=["blue", "green"])  
  
plt.title("Heart Disease values")  
  
plt.xlabel("1 Heart Disease, No Heart Disease")  
  
plt.ylabel("Amount")  
  
plt.show()
```



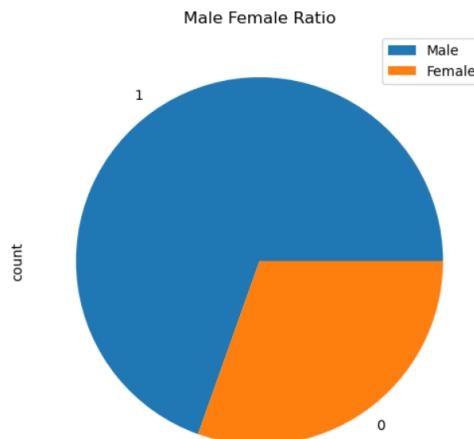
```
[20]: import matplotlib.pyplot as plt  
  
df.target.value_counts().plot(kind='pie', figsize=(8, 6))  
  
plt.legend(["Disease", "No Disease"])  
plt.show()
```



```
[22]: df.sex.value_counts()
```

```
[22]: sex  
1    713  
0    312  
Name: count, dtype: int64
```

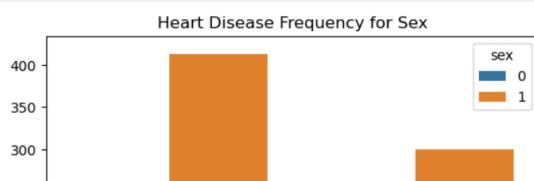
```
[23]: import matplotlib.pyplot as plt  
  
df.sex.value_counts().plot(kind='pie', figsize=(8, 6))  
plt.title('Male Female Ratio')  
plt.legend(['Male', 'Female'])  
plt.show()
```

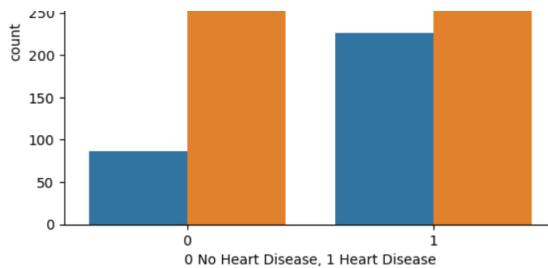


```
[25]: #question 2  
import pandas as pd  
  
cross_tab = pd.crosstab(df.target, df.sex)  
print(cross_tab)
```

```
sex      0   1  
target  
0        86  413  
1       226  300
```

```
[26]: import seaborn as sns  
  
import matplotlib.pyplot as plt  
  
sns.countplot(x='target', data=df, hue='sex')  
  
plt.title('Heart Disease Frequency for Sex')  
  
plt.xlabel('0 No Heart Disease, 1 Heart Disease')  
  
plt.show()
```



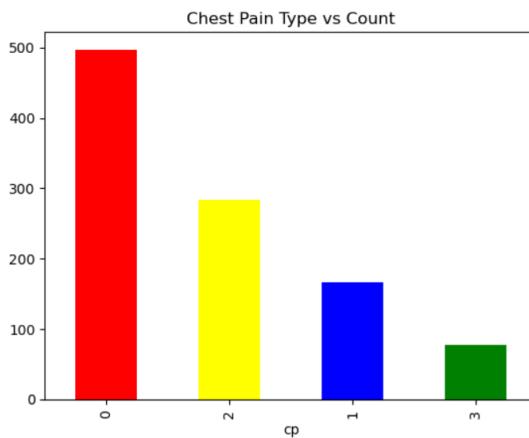


```
[27]: #question 3
df.cp.value_counts()
```

```
[27]: cp
0    497
2    284
1    167
3     77
Name: count, dtype: int64
```

```
[32]: import matplotlib.pyplot as plt

# Assuming 'df' is your DataFrame and 'cp' is the column name
df.cp.value_counts().plot(kind='bar', color=["red", "yellow", "blue", "green"])
plt.title('Chest Pain Type vs Count')
plt.show()
```



```
[5]: import pandas as pd
df = pd.read_csv('C:\\Users\\siva\\Downloads\\heart.csv')
crosstab_result = pd.crosstab(df.sex, df.cp)
print(crosstab_result)
```

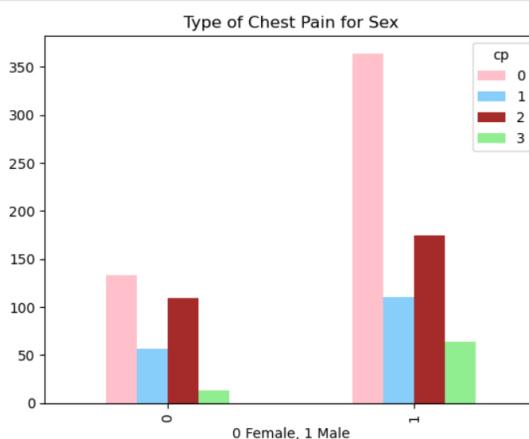
```
cp    0    1    2    3
sex
0    133   57  109   13
1    364  110  175   64
```

```
[7]: import pandas as pd
import matplotlib.pyplot as plt

pd.crosstab(df.sex, df.cp).plot(kind='bar', color=["pink", "lightskyblue", "brown", "lightgreen"])

plt.title('Type of Chest Pain for Sex') # Corrected quotation marks
plt.xlabel('0 Female, 1 Male') # Corrected quotation marks

plt.show()
```



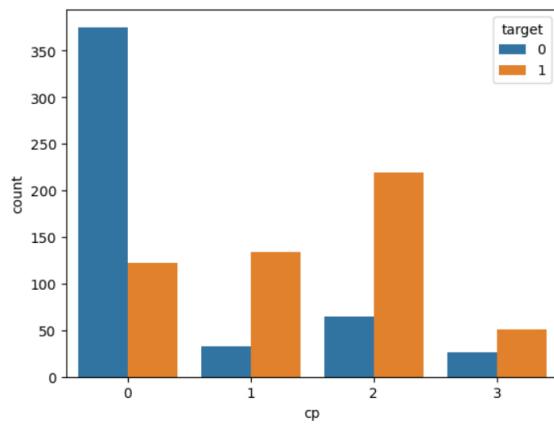
```
[8]: #question 4
```

```
pd.crosstab(df.cp, df.target)
```

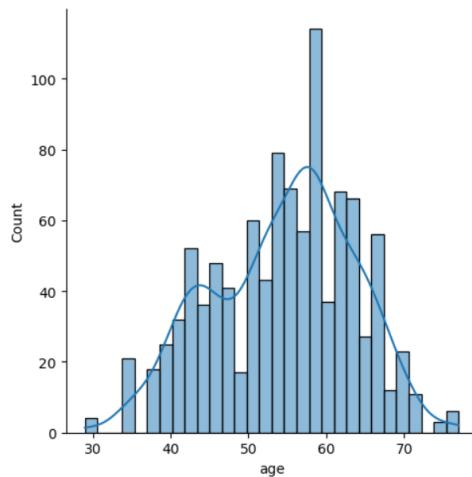
```
[8]: target    0    1
cp
```

```
0 375 122  
1 33 134  
2 65 219  
3 26 51
```

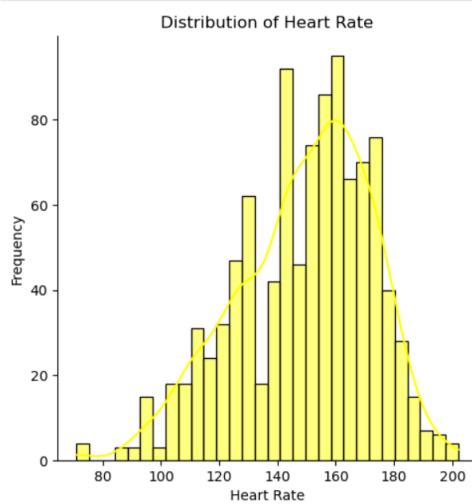
```
[11]: import seaborn as sns  
sns.countplot(x = 'cp', data = df, hue='target');
```



```
[16]: import seaborn as sns  
sns.distplot(x='age', data=df, bins=30, kde=True);
```



```
[17]: import seaborn as sns  
import matplotlib.pyplot as plt  
  
# Assuming 'df' is your DataFrame and 'thalach' is the column for heart rate  
sns.distplot(x='thalach', data=df, bins=30, kde=True, color='yellow')  
  
plt.title('Distribution of Heart Rate')  
plt.xlabel('Heart Rate')  
plt.ylabel('Frequency')  
plt.show()
```



```
[21]: import pandas as pd  
  
# Assuming 'df' is your DataFrame
```

```
cross_tab = pd.crosstab(df.cp, df.target)
```

```
print(cross_tab)
```

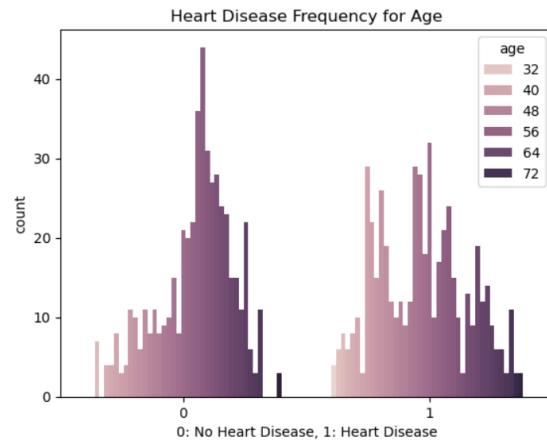
```
target    0    1
cp
0        375  122
1         33   134
2         65   219
3         26   51
```

```
[23]: import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
sns.countplot(x='target', data=df, hue='age')
```

```
plt.title('Heart Disease Frequency for Age')
plt.xlabel('0: No Heart Disease, 1: Heart Disease')
plt.show()
```



```
[24]: #question 6
df.cp.value_counts()
```

```
[24]: cp
0    497
2    284
1    167
3     77
Name: count, dtype: int64
```

```
[26]: pd.crosstab(df.age, df.target)
```

```
[26]: target    0    1
```

```
age
```

```
29    0    4
```

```
34    0    6
```

```
35    7    8
```

```
37    0    6
```

```
38    4    8
```

```
39    4   10
```

```
40    8    3
```

```
41    3   29
```

```
42    4   22
```

```
43   11   15
```

```
44   10   26
```

```
45    6   19
```

```
46   11   12
```

```
47    8   10
```

```
48   11   12
```

```
49    8    9
```

```
50    9   12
```

```
51   10   29
```

```
52   15   28
```

```
53    8   18
```

```
54   21   32
```

```
55   20   10
```

```
56   22   17
```

```
57   36   21
```

```
58   44   24
```

```
59   31   15
```

```
60   27   10
```

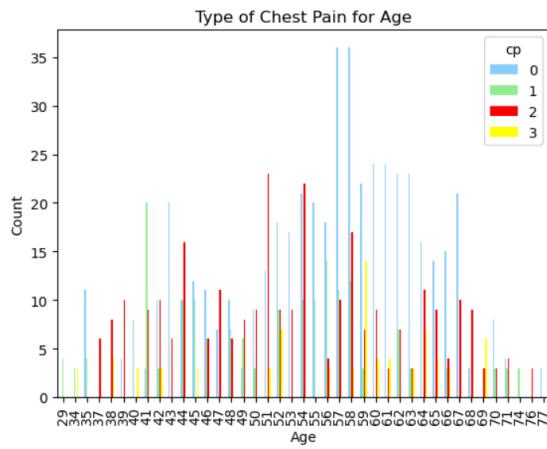
```
61   28    3
```

```
62   24   13
```

```
63   23    9
```

```
64 15 19  
65 15 12  
66 11 14  
67 22 9  
68 6 6  
69 3 6  
70 11 3  
71 0 11  
74 0 3  
76 0 3  
77 3 0
```

```
[27]: import pandas as pd  
import matplotlib.pyplot as plt  
  
# Assuming df is already defined  
pd.crosstab(df.age, df.cp).plot(kind='bar', color=["lightskyblue", "lightgreen", "red", "yellow"])  
  
plt.title('Type of Chest Pain for Age')  
plt.xlabel('Age')  
plt.ylabel('Count')  
plt.show()
```



```
[ ]: 
```